Hepatoprotective effects of systemic ER activation

Patient cohort analysis - determining markers to separate early and advanced liver disease

Christian Sommerauer & Carlos Gallardo

15 August, 2022

```
library(tidyverse)
```

Load the fully annotated table and filter the relevant 38 genes. Create a dataframe with all annotations.

```
cohort.rds <- readRDS("/Users/christian.som/Documents/GitHub/MAFLD_ER_agonists/data/bulkRNAseq_human_co
my_38genes <- scan('results/ER_regulated_genes.txt', what = character(), quiet = T)

mouse_human_orthologs <- read.table(
  file = 'data/ensembl_mmus_hsap_sep2019_orthologs.tsv',
  sep = '\t',
  header = TRUE,
  quote = '')

cohort.rds$Govaere$cpm_filt <- cohort.rds$Govaere$cpm %>%
  tibble::rownames_to_column(var = 'gene') %>%
  dplyr::filter(gene %in% mouse_human_orthologs$GeneID_human) %>%
  dplyr::mutate(gene = dplyr::recode(gene, !!!setNames(mouse_human_orthologs$GeneSymbol_human,
                                                       mouse_human_orthologs$GeneID_human))) %>%
  dplyr::filter(!duplicated(gene) & gene != "") %>%
  tibble::column_to_rownames(var = 'gene')

df <- cohort.rds$Govaere$cpm_filt %>%
  t() %>%
  as.data.frame() %>%
  tibble::rownames_to_column(var="Patient") %>%
  pivot_longer(cols=2:ncol(.), names_to = "gene") %>%
  inner_join(cohort.rds$Govaere$meta)
```

```
loop.me <- my_38genes

#Make lists for the summary statistics
res.aov_stage_list <- list()
res.aov_NAS_list <- list()
res.aov_fib_list <- list()
```

```
for (i in 1:38) {

#Group
object.1 <- df %>% filter(gene==c(loop.me[i])) %>% group_by(Fibrosis)

# Compute the analysis of variance
res.aov_fib <- aov(value ~ Fibrosis, data = object.1)
# Summary of the analysis
summary(res.aov_fib)
# Multiple pairwise comparisons
res.aov_fib <- TukeyHSD(res.aov_fib)
res.aov_fib_list[[i]] <- as.data.frame(res.aov_fib$Fibrosis)

#Group
object1.plot.NAS <- object.1 %>% group_by(NAS)
# Compute the analysis of variance
res.aov_NAS <- aov(value ~ NAS, data = object1.plot.NAS)
# Summary of the analysis
summary(res.aov_NAS)
# Multiple pairwise comparisons
res.aov_NAS <- TukeyHSD(res.aov_NAS)

res.aov_NAS_list[[i]] <- as.data.frame(res.aov_NAS$NAS)

#Group
object1.plot.stage <- object.1 %>% group_by(Stage)
# Compute the analysis of variance
res.aov_stage <- aov(value ~ Stage, data = object1.plot.stage)
# Summary of the analysis
summary(res.aov_stage)
# Multiple pairwise comparisons
res.aov_stage <- TukeyHSD(res.aov_stage)

res.aov_stage_list[[i]] <- as.data.frame(res.aov_stage$Stage)

}
```

Filter for significant p-values among comparisons.[Stage]

```
names(res.aov_stage_list) <- loop.me

res.aov_stage_list_2 <- list()
for (i in 1:length(res.aov_stage_list)) {
  res.aov_stage_list_2[[i]] <- res.aov_stage_list[[i]] %>% tibble::rownames_to_column("contrast")
  res.aov_stage_list_2[[i]]$Gene_symbol <- names(res.aov_stage_list[i])
}

res.aov_stage_list_df <- do.call(rbind, unname(res.aov_stage_list_2))
res.aov_stage_list_df_sign <- res.aov_stage_list_df %>% rename("p"= "p adj") %>% filter(p<0.01)
```

Filter for significant p-values among comparisons [Fibrosis]

```
names(res.aov_fib_list) <- loop.me

res.aov_fib_list_2 <- list()
for (i in 1:length(res.aov_fib_list)) {
  res.aov_fib_list_2[[i]] <- res.aov_fib_list[[i]] %>% tibble::rownames_to_column("contrast")
  res.aov_fib_list_2[[i]]$Gene_symbol <- names(res.aov_fib_list[i])
}

res.aov_fib_list_df <- do.call(rbind, unname(res.aov_fib_list_2))
res.aov_fib_list_df_sign <- res.aov_fib_list_df %>% rename("p"= "p adj") %>% filter(p<0.01)
```

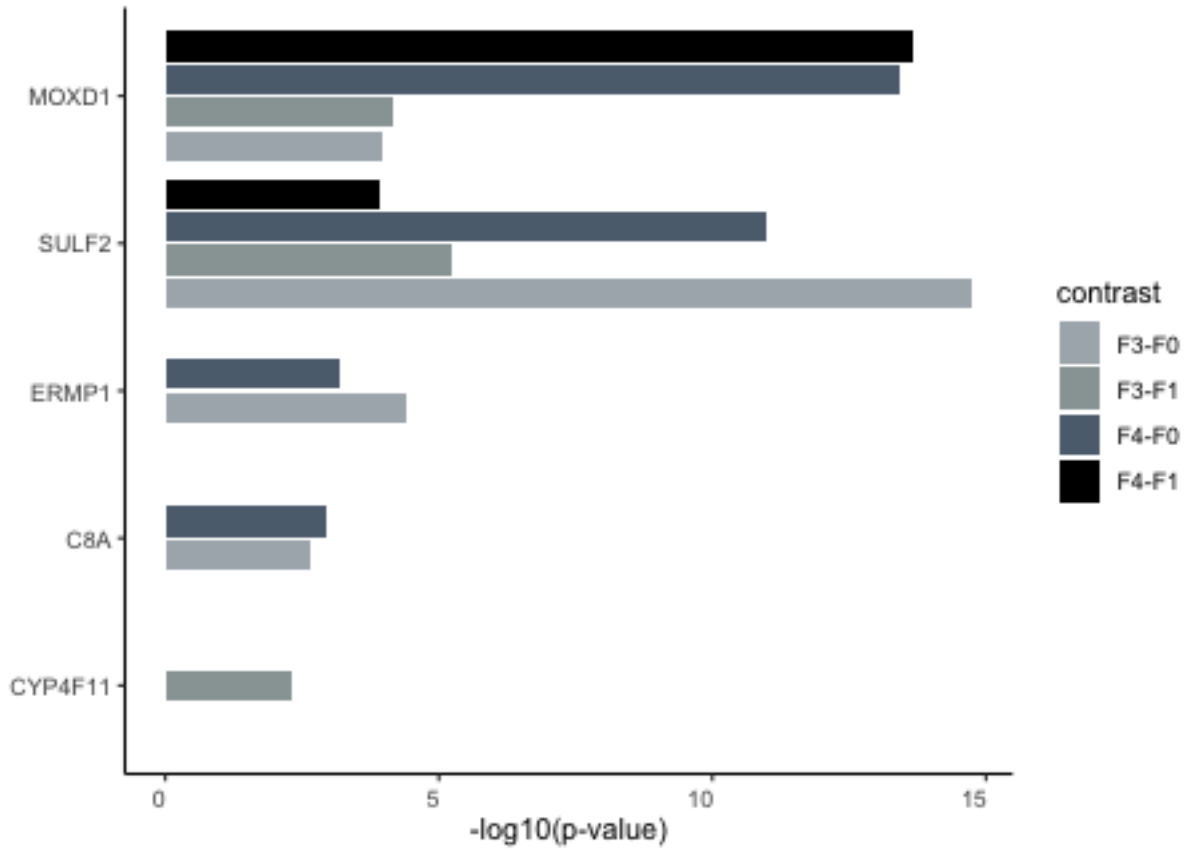
Filter for significant p-values among comparisons [NAS]

```
names(res.aov_NAS_list) <- loop.me

res.aov_NAS_list_2 <- list()
for (i in 1:length(res.aov_NAS_list)) {
  res.aov_NAS_list_2[[i]] <- res.aov_NAS_list[[i]] %>% tibble::rownames_to_column("contrast")
  res.aov_NAS_list_2[[i]]$Gene_symbol <- names(res.aov_NAS_list[i])
}

res.aov_NAS_list_df <- do.call(rbind, res.aov_NAS_list_2)
res.aov_NAS_list_df_sign <- res.aov_NAS_list_df %>% rename("p"= "p adj") %>% filter(p<0.01)
```

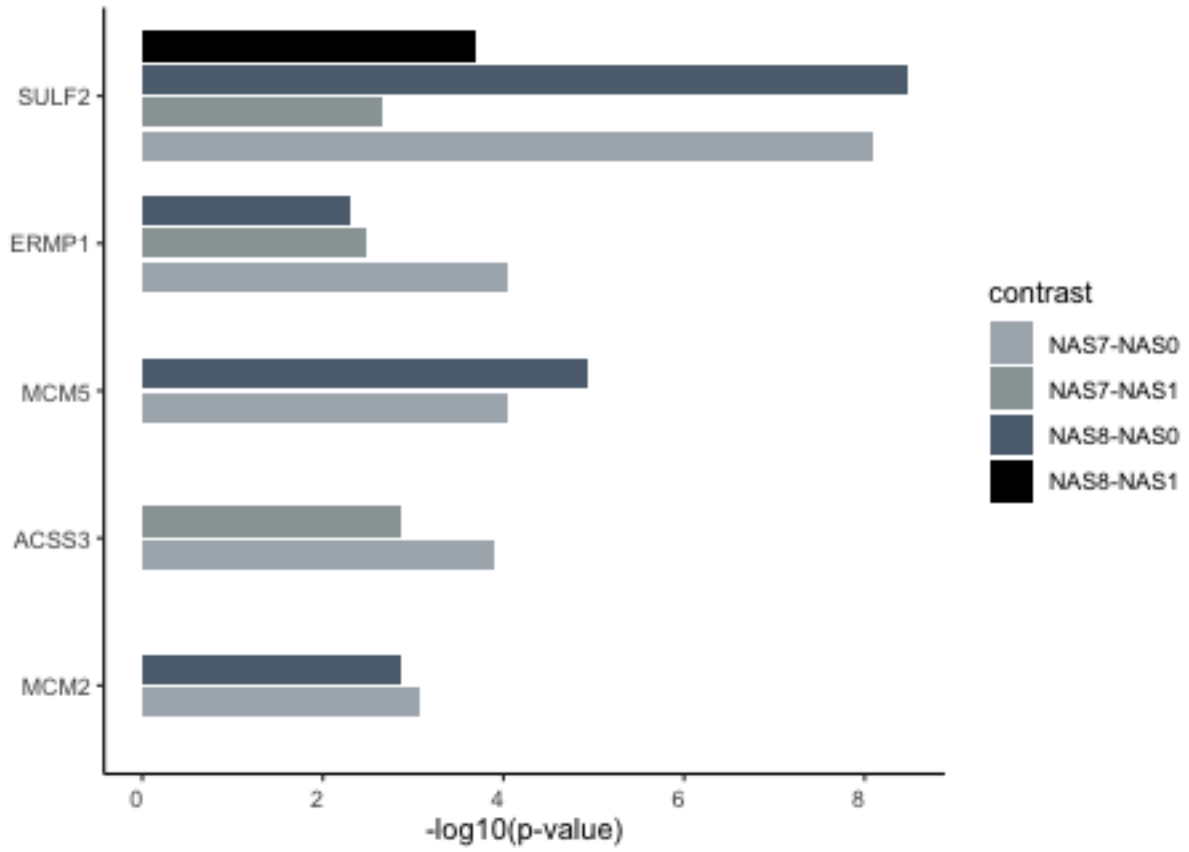
```
stage.anova <- res.aov_stage_list_df_sign
fib.anova <- res.aov_fib_list_df_sign
NAS.anova <- res.aov_NAS_list_df_sign
```

```
# These genes are have conserved gene expression trends from mouse to
# human and are not changed in female HFD mice (Heatmap Fig6A)
ER_regulated_genes_conserved21genes <- scan("/Users/christian.som/Documents/GitHub/MAFLD_ER_agonists/res
```

```
fib.anova.2 <- fib.anova %>% dplyr::select(-diff, -lwr, -upr) %>%
  filter(Gene_symbol %in% ER_regulated_genes_conserved21genes) %>%
  filter(contrast %in% c("F3-F0", "F4-F0", "F3-F1", "F4-F1"))

fib.anova.order <- fib.anova.2 %>% add_count(Gene_symbol, sort = TRUE) %>%
dplyr::pull(Gene_symbol) %>% unique()

# BAR PLOT
ggplot(fib.anova.2, aes(x=-log10(p), y=factor(Gene_symbol, levels=rev(fib.anova.order)))) +
  geom_col(aes(fill=contrast), position = position_dodge2(preserve = "single")) +
  theme_classic() +
  theme(axis.text.x = element_text(hjust=0.95, vjust=0.5)) +
  ylab("") +
  xlab("-log10(p-value)") +
  scale_fill_manual(values=c("#abb2b9", "#99a3a4", "#5d6d7e", "#000000"))
```




```
NAS.anova.2 <- NAS.anova %>% dplyr::select(-diff, -lwr, -upr) %>%
  filter(Gene_symbol %in% ER_regulated_genes_conserved21genes) %>%
  filter(contrast %in% c("NAS8-NAS0", "NAS8-NAS1", "NAS7-NAS0", "NAS7-NAS1"))
NAS.anova.order <- NAS.anova.2 %>% add_count(Gene_symbol, sort = TRUE) %>%
dplyr::pull(Gene_symbol) %>% unique()

# BAR PLOT
ggplot(NAS.anova.2, aes(x=-log10(p), y=factor(Gene_symbol, levels=rev(NAS.anova.order)))) +
  geom_col(aes(fill=contrast), position = position_dodge2(preserve = "single")) +
  theme_classic() +
  theme(axis.text.x = element_text(hjust=0.95, vjust=0.5)) +
  ylab("") +
  xlab("-log10(p-value)") +
  scale_fill_manual(values=c("#abb2b9", "#99a3a4", "#5d6d7e", "#000000"))
```



Export names of the marker genes

```
fib_markers <- unique(fib.anova.2$Gene_symbol)
NAS_markers <- unique(NAS.anova.2$Gene_symbol)

cat(fib_markers, file="results/ER_regulated_genes_fibrosis_markers.txt", sep="\n")
cat(NAS_markers,  file="results/ER_regulated_genes_NAS_markers.txt", sep="\n")
```

```
sessionInfo()
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] forcats_0.5.1   stringr_1.4.0   dplyr_1.0.6     purrr_0.3.4
```

```
## [5] readr_1.4.0     tidyr_1.1.3     tibble_3.1.2    ggplot2_3.3.3
## [9] tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.1.1  xfun_0.31         haven_2.4.1       colorspace_2.0-1
##  [5] vctrs_0.3.8       generics_0.1.0    htmltools_0.5.1.1 yaml_2.2.1
##  [9] utf8_1.2.1        rlang_0.4.11      pillar_1.6.1      glue_1.6.2
## [13] withr_2.4.2       DBI_1.1.1         dbplyr_2.1.1      modelr_0.1.8
## [17] readxl_1.3.1      lifecycle_1.0.0   munsell_0.5.0     gtable_0.3.0
## [21] cellranger_1.1.0  rvest_1.0.0       evaluate_0.14     labeling_0.4.2
## [25] knitr_1.33        fansi_0.5.0       highr_0.9         broom_0.7.6
## [29] Rcpp_1.0.6        scales_1.1.1      backports_1.2.1   jsonlite_1.7.2
## [33] farver_2.1.0      fs_1.5.0          hms_1.1.0         digest_0.6.27
## [37] stringi_1.6.2     grid_4.0.3        cli_3.2.0         tools_4.0.3
## [41] magrittr_2.0.1    crayon_1.4.1      pkgconfig_2.0.3   ellipsis_0.3.2
## [45] xml2_1.3.2        reprex_2.0.0      lubridate_1.7.10  assertthat_0.2.1
## [49] rmarkdown_2.14    httr_1.4.2        rstudioapi_0.13   R6_2.5.0
## [53] compiler_4.0.3
```